Statistics-based filtering for low signal-to-noise ratios, applied to rocket plume imaging

Harald Hovland
Norwegian Defence Research Establishment, Postboks 25, N-2027 Kjeller, NORWAY

ABSTRACT

Extracting information from low signal to noise ratio images poses significant challenges. Noise makes extracting spatial features difficult, in particular if extraction of both large, smooth features at the same time as point-like features is required. This work describes a new statistical approach, able to handle both simultaneously, with the capacity of handling both positive and negative contrast signatures. The basic idea in this approach is that each pixel value can represent underlying statistics to a varying degree, depending on how similar it is to samples taken close to it, spatially and/or temporally. If the sample is similar to its surroundings, it is strongly filtered and also affects the filtering of neighboring samples, but if it is significantly different, it will remain largely unfiltered and does not influence neighboring pixel filtering. Simulations show that the filtering maintains energy conservation, significantly limits noise and at the same time maintains signal integrity. The filter is found to adapt to noise characteristics and spatiotemporal variations of the background. The technique is found to be well suited to rocket plume imaging, but is adaptable to a broad range of other applications.

Keywords: Feature extraction, imaging, probability density distribution, adaptive filtering, ultraviolet, infrared, noise suppression, CFAR.

1. INTRODUCTION

Detecting missile plumes at long range distance with an ultraviolet (UV) or infrared (IR) imager is a typical case of detecting small sources with relatively small signal-to-noise ratios. In the solar blind UV region[1], the number of very energetic photons arriving at the detector is likely to be small, and the source will typically have a small extent. Noise in the imager can then easily make the detection process a challenge. Likewise, in the infrared portion of the spectrum, clutter will typically be significant compared to the source, which could easily be of subpixel size. Even with low noise readout, background noise could still be challenging. An incoming missile will also increase its angular extent as seen from the detector as it approaches. Another challenge is non-uniformity in both background level and noise.

A classical way to remove clutter is to apply a low-pass filter to the image, but this will normally destroy point sources and decimate small area sources.

In modern low photon count imaging[2], notably ghost imaging and heralded imaging, a photoelectron count at a detector is not simply taken as being direct proportional to the source output, but rather as a sampling of an underlying distribution. Hence, a readout of 12 photoelectrons at an image pixel may represent a sample of a Poisson distribution with a mean corresponding approximatively to between 9 and 16 photoelectrons.

We here propose a local filtering process in which signals are low pass filtered only if they are statistically similar to the local level, but remains largely unfiltered if they differ sufficiently from the background. They will also contribute less to the statistics of the local filter the less they resemble the local level.

2. THEORY

2.1 Introduction of adherence statistics

In the proposed approach, the concept of adherence is introduced to express how well a signal sampled at a pixel blends in with those of the surrounding pixels. This concept bears a certain resemblance to the Constant False Alarm Rate (CFAR) approach[3]. The surrounding signals can be samples belonging to pixels that are spatially and/or temporally close

to the pixel under investigation. The adherence is a measure of how well individual sample values are associated with the local (statistical) distribution of the pixels surrounding it spatially and/or temporally.

It is thus a requirement to have some idea of the shape or nature of the distribution, without setting the characteristic parameters beforehand. For example, one might assume the distribution to be Gaussian in nature. This kind of assumption can often be justified by the physical nature of the scene. As an example, the output statistics of a source near thermal equilibrium would normally be close to a Poisson distribution, which at higher photon counts is well approximated to a Gaussian distribution.

We now consider the local distribution of a pixel *j*. Formally, the sample value $x_{i,t}$ at time *t* of another pixel *i* is considered to have an **adherence** $a_{i,j,t}$ to the distribution local to pixel *j*, its adherence depending on its similarity to the distribution. The adherence is a function of time *t* due to the fact that the samples vary over time. The local distribution of pixel *j* is calculated by the point distribution, weighted by their adherence, such that one can calculate the mean $\mu_{j,t}$, variance $\sigma_{j,t}^2$ or a more general statistical property $\langle f(x)\rangle_{j,t}$:

$$\mu_{j,t} = \langle x \rangle_{j,t} = \frac{\sum_i x_{i,t} a_{i,j,t}}{\sum_i a_{i,j,t}} \tag{1}$$

$$\sigma_{j,t}{}^2 = \left\langle \left(x - \mu_{j,t}\right)^2 \right\rangle_{j,t} = \left\langle x^2 - 2x\mu_{j,t} + \mu_{j,t}{}^2 \right\rangle_{j,t} = \frac{\sum_i x_{i,t}{}^2 a_{i,j,t}}{\sum_i a_{i,j,t}} - \mu_{j,t}{}^2 \tag{2}$$

$$\langle f(x) \rangle_{j,t} = \frac{\sum_i f\left(x_{i,t}\right) a_{i,j,t}}{\sum_i a_{i,j,t}} \tag{3}$$

The adherence is defined as follows:

$$a_{i,j,t} = \sqrt{\frac{p_{j,t}\left(x_{i,t}\right)}{\max\left(p_{j,t}\right)}} \tag{4}$$

Here $p_{j,t}(x_{i,t})$ is the probability density of the distribution local to pixel *j* at the sample pixel *i* at time *t*, and $\max(p_{j,t})$ is the maximum value of this probability density. That this choice makes sense will be argued for later. In the case where the local distribution is Gaussian (an important case because of the central limit theorem), we have:

$$a_{i,j,t} = \sqrt{e^{-\frac{\left(x_{i,t}-\mu_{j,t}\right)^2}{2\sigma_{j,t}{}^2}}} = e^{-\frac{\left(x_{i,t}-\mu_{j,t}\right)^2}{4\sigma_{j,t}{}^2}}, \quad \mu_{j,t} = \frac{\sum_i x_{i,t} a_{i,j,t}}{\sum_i a_{i,j,t}}, \quad \sigma_{j,t}{}^2 = \frac{\sum_i \left(x_{i,t} - \mu_{j,t}\right)^2 a_{i,j,t}}{\sum_i a_{i,j,t}} \tag{5}$$

The expression for $a_{i,j}$ can then be found iteratively by inserting for μ_j and $\sigma_j{}^2$ in (5):

$$a_{i,j,t} = e^{-\frac{\left(x_{i,t}-\mu_{j,t}\right)^2}{4\frac{\left(\sum_k \left(x_{k,t}-\mu_{j,t}\right)^2 a_{k,j,t}\right)^2}{\sum_k a_{k,j,t}}}} = e^{-\frac{\left(\sum_k \left(x_{i,t}-x_{k,t}\right) a_{k,j,t}\right)^2}{4\left(\left(\sum_k a_{k,j,t} \sum_k x_{k,t}{}^2 a_{k,j,t}\right) - \left(\sum_k x_{k,t} a_{k,j,t}\right)^2\right)}} \tag{6}$$

Initially inserting $a_{i,j,t}$=1 on the right hand side of the expression, and then take the resulting left hand side expression of $a_{i,j,t}$ and insert it on the right hand side again iteratively, the expression for $a_{i,j,t}$ is found to converge rather quickly (12 iterations could give an approximate 10^{-5} relative error, depending on the statistics).

Once the adherence coefficients $a_{i,j,t}$ have been determined, a filter correction can be made to the sample value. In the case of a Gaussian distribution, we choose:

$$X_{i,t} = f_{filter,i,t}\left(x_{i,t}\right) = \mu_{i,t} {a_{i,i,t}}^2 + x_{i,t}\left(1 - {a_{i,i,t}}^2\right) \tag{7}$$

That this makes sense can be argued for in the case of a Gaussian distribution. In the unfiltered case, we would have:

$$X_{i,t,unfiltered} = x_{i,t}, \tag{8}$$

which is the case when $a_{i,i,t}=0$ (i.e., no adherence). At $x_{i,t}=\mu$, the distribution is at its maximum, and we have:

$$X_{i,t} = \mu_{i,t} {a_{i,i,t}}^2 + x_{i,t}\left(1 - {a_{i,i,t}}^2\right) = \mu_{i,t}\left(\frac{p\left(\mu_{i,t}\right)}{\max\left(p_{i,t}\right)}\right) + x_{i,t}\left(1 - \left(\frac{p\left(\mu_{i,t}\right)}{\max\left(p_{i,t}\right)}\right)\right) = \mu_{i,t}, \tag{9}$$

The derivative of the function with respect to $x_{i,t}$ is given as:

$$\begin{aligned}
&\frac{d}{dx_{i,t}} X_{i,t} = \frac{d}{dx_{i,t}}\left(\mu_{i,t} {a_{i,i,t}}^2 + x_{i,t}\left(1 - {a_{i,i,t}}^2\right)\right) \\
&= \mu_{i,t} \frac{d}{dx_{i,t}}\left(\frac{p\left(x_{i,t}\right)}{\max\left(p_{i,t}\right)}\right) + \left(1 - \left(\frac{p\left(x_{i,t}\right)}{\max\left(p_{i,t}\right)}\right)\right) - x_{i,t} \frac{d}{dx_{i,t}}\left(\frac{p\left(x_{i,t}\right)}{\max\left(p_{i,t}\right)}\right), \\
&= \left(1 - \left(\frac{p\left(x_{i,t}\right)}{\max\left(p_{i,t}\right)}\right)\right) + \left(\mu_{i,t} - x_{i,t}\right) \frac{d}{dx_{i,t}}\left(\frac{p\left(x_{i,t}\right)}{\max\left(p_{i,t}\right)}\right)
\end{aligned} \tag{10}$$

The expression in (10) equals 0 at the peak probability, as expected.

The choice of samples to include in the distribution local to pixel *j* has not yet been specified, but it is natural to consider spatio-temporal neighbors of a given pixel sample. In practice, the choice of neighbors should be based on physical aspects around the detection process, such as atmospheric scattering, solar irradiation and scattering, and camera/sensor properties. The computational capacity available must evidently also be taken into account in a practical implementation.

It is possible to generalize the choice of samples by including them as a factor in a **generalized adherence coefficient *h***:

$$h_{i,j,t,t'} = a_{i,j,t,t'} g_{i,j,t,t'} \tag{11}$$

Here, $g_{i,j,t,t'}$ is a spatio-temporal factor describing the connection between samples taken at different pixels (*i* and *j*) and different times (*t* and *t'*). The adherence would then be generalized to:

$$a_{i,j,t,t'} = \sqrt{\frac{p_{j,t}\left(x_{i,t'}\right)}{\max\left(p_{j,t}\right)}} \tag{12}$$

The general case of adherence statistics then becomes:

$$\mu_{j,t} = \left\langle x \right\rangle_{j,t} = \frac{\sum_{i,t'} x_{i,t'} h_{i,j,t,t'}}{\sum_{i,t'} h_{i,j,t,t'}} = \frac{\sum_{i,t'} x_{i,t'} a_{i,j,t,t'} g_{i,j,t,t'}}{\sum_{i,t'} a_{i,j,t,t'} g_{i,j,t,t'}} \tag{13}$$

$$\sigma_{j,t}{}^2 = \left\langle \left(x-\mu_{j,t}\right)^2 \right\rangle_{j,t} = \left\langle x^2 - 2x\mu_{j,t} + \mu_{j,t}{}^2 \right\rangle_{j,t} = \frac{\sum_{i,t'} x_{i,t'}{}^2 h_{i,j,t,t'}}{\sum_{i,t'} h_{i,j,t,t'}} - \mu_{j,t1}{}^2 = \frac{\sum_{i,t'} x_{i,t'}{}^2 a_{i,j,t,t'} g_{i,j,t,t'}}{\sum_{i,t'} a_{i,j,t,t'} g_{i,j,t,t'}} - \mu_{j,t}{}^2 \quad (14)$$

$$\left\langle f\left(x\right)\right\rangle_{j,t1} = \frac{\sum_{i,t'} f\left(x_{i'}\right) h_{i,j,t,t'}}{\sum_{i,t'} h_{i,j,t,t'}} = \frac{\sum_{i,t'} f\left(x_{i'}\right) a_{i,j,t,t'} g_{i,j,t,t'}}{\sum_{i,t'} a_{i,j,t,t'} g_{i,j,t,t'}} \quad (15)$$

It is important to note that in principle, the sums here are taken over all pixels and samples at all times. The filtered signal would then become:

$$X_{i,t} = f_{filter,i,t}\left(x_{i,t}\right) = \mu_{i,t} h_{i,i,t,t}{}^2 + x_{i,t}\left(1-h_{i,i,t,t}{}^2\right) = \mu_{i,t} a_{i,i,t,t}{}^2 g_{i,i,t,t}{}^2 + x_{i,t}\left(1-a_{i,i,t,t}{}^2 g_{i,i,t,t}{}^2\right) \quad (16)$$

Assuming the spatio-temporal factor $g_{i,j,t,t'}$ is unity for identical samples ($i=j$) and times ($t=t'$), (16) reduces to:

$$X_{i,t} = f_{filter,i,t}\left(x_{i,t}\right) = \mu_{i,t} a_{i,i,t,t}{}^2 + x_{i,t}\left(1-a_{i,i,t,t}{}^2\right) \quad (17)$$

The geometric factor $g_{i,j,t,t'}$ would in principle be a four-dimensional matrix (assuming discrete spatial and temporal samples). Its actual form would be chosen according to the physics of the source and background, atmosphere and sensor electrooptics. It is not unreasonable to assume that the dimensionality could be reduced, both to a time difference $(t,t') \rightarrow \Delta t = t-t'$, and, if the point spread function of the sensor is uniform, to a distance $d_{i,j}$ between samples i and j. We then get:

$$g_{i,j,t,t'} = g_{d_{i,j},\Delta t} \quad (18)$$

2.2 Multimodal case

An issue that can be noted is that the model presented in section 2.1 is expected to behave badly if the local background consists of several components which differ significantly in nature, for example when the scene consists of both land and sea. In this case, the model breaks down in the transition area because statistical properties like the mean and variance does not represent well the nature of the variations.

3. SIMULATIONS

3.1 Basic filter behavior

As a starting point, 4000 samples were simulated with a Gaussian distribution with a mean of 5 and a standard deviation of 1, and 100 samples were given a regularly distributed value from 0 to 10.

The adherence levels a_i were initially all set to 1. Using these values, the filter output values for all input values, become the mean value of the distribution. This is shown in figure 1 (blue curve). Using an iterative determination of a_i, the values are found to quickly converge (red curves), resulting in the green curve. It is seen that for values close to the distribution mean, a hard filtering is applied, whereas for sample values far away from the mean value (relative to the standard deviation of the curve), the values are almost unfiltered. The adherence statistics filter (green) curve is found to move with the statistical mean value and scale with the standard deviation.

The integrated sum of output filtered values is found to have the same value as the integrated sum of input values, indicating that the method conserves energy well. This is also found for more asymmetric variations. The reason for this is that when a value deviates significantly from the mean, it is only influenced by, or influences the other values to a very small degree.

In the case of real-time calculations, the temporal aspect could imply an asymmetry in the fact that future signals will not affect past signals (because they are not available at the time), whereas past signals could affect future ones. Purely spatially separated signals will normally have a more symmetrical influence on each other.

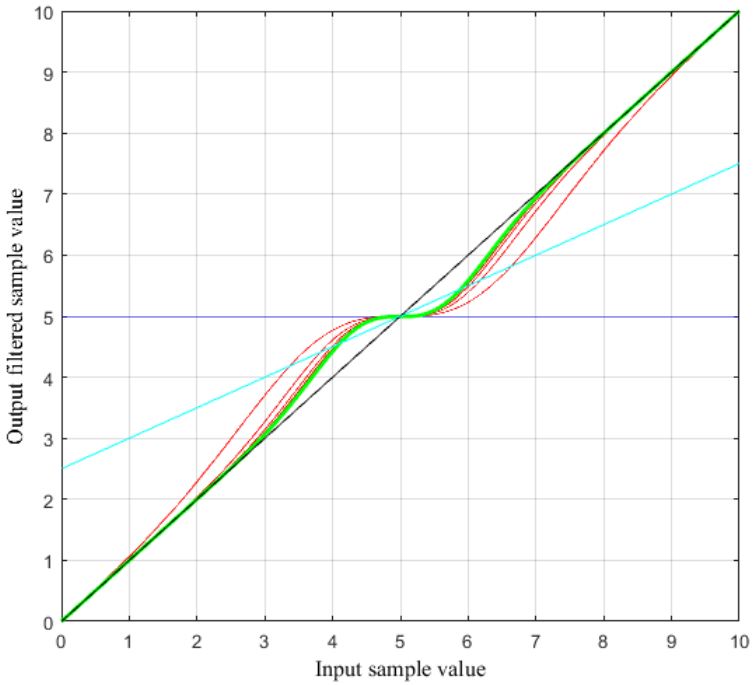

Figure 1. Filtered sample value as a function of input. The sample values consists of 4000 sample values with a Gaussian distribution, with a mean value of 5 and a standard deviation of 1, and 128 sample points regularly distributed from 0 to 10. With the start values of the adherence factors *a* set to 1, all sample values are filtered to the mean value (blue curve). The adherence factors are then calculated iteratively, producing the red curves, until the values of *a* converge, resulting in the filter output represented by the green curve. It can be seen that for values close to the mean of the statistics, the sample values are strongly filtered, whereas for values far away from the mean value, the sample values are almost unfiltered. The black curve corresponds to an unfiltered output, and the magenta curve corresponds to the halfway between the completely filtered (blue curve) and the unfiltered (black curve) levels.

3.2 Spatial filtering

Assuming the hardware is optimized, it should be possible to achieve background noise levels at the sensor output. A 250×550 pixel static scene was generated with a linear ramp background signal level change across the scene with zero photoelectrons per pixel at the 51 leftmost columns side and 650 photoelectrons (RMS) per pixel at the right side (and hence a square root noise level profile, with 0 photoelectrons RMS noise per pixel at the left side and 25 photoelectrons RMS noise per pixel at the right side). In the analysis, a 50 pixel wide border was removed in the images to avoid any edge effects due to the windowing in the filtering process. (A 101×101 pixel window is used in the filtering process).

6 sources were superposed on the scene. 3 single (sub)pixel sources producing 100 photoelectrons in addition to the background were located in x=100, 200, 300, and y=150, and 3 extended Gaussian sources, each subtending approximately 172 pixels (max/e²) and producing approximately 7.4×10^3 photoelectrons, with an average peak intensity of approximately 93 photoelectrons per pixel. The extended sources were centered on x=100, 200, 300, and y=100. Figure 2 and 3 shows the resulting image without and with a linear background subtraction. There is significant noise present in the image.

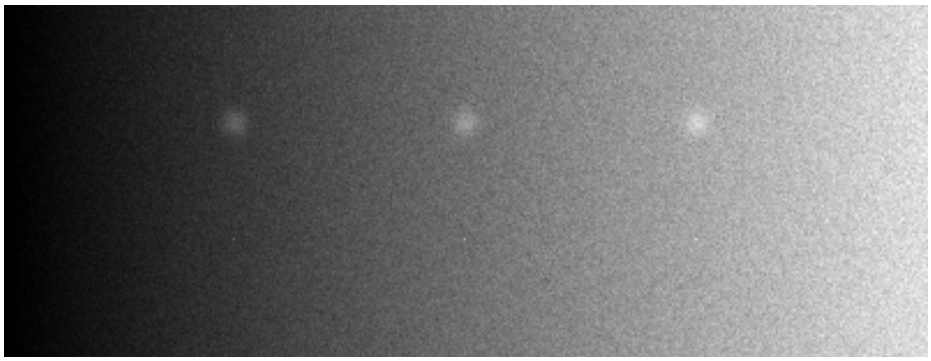

Figure 2. Gradient background image with 3 point sources and 3 extended sources. The background noise is visible.

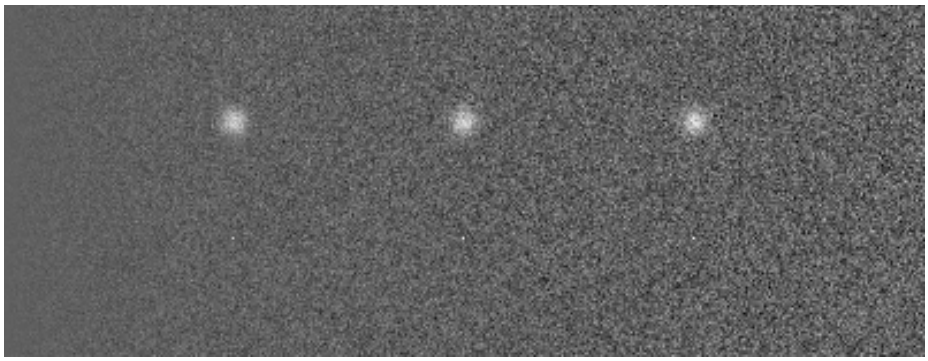

Figure 3. Image in figure 2, with a linear background subtraction. The background noise is significant.

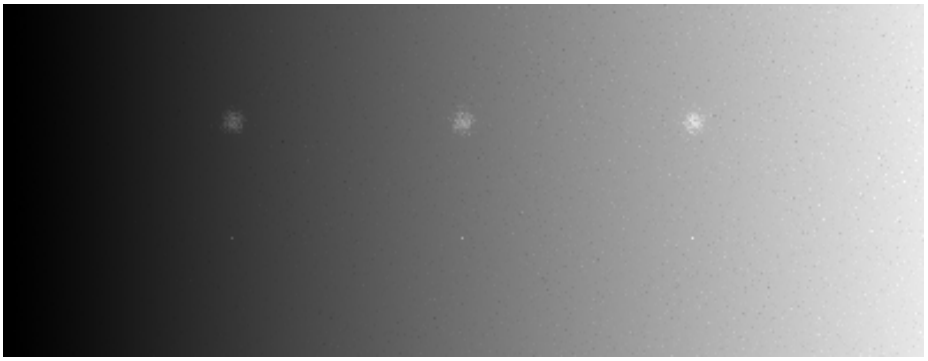

Figure 4. Adherence statistics filtered version of figure 2. The background noise is significantly suppressed.

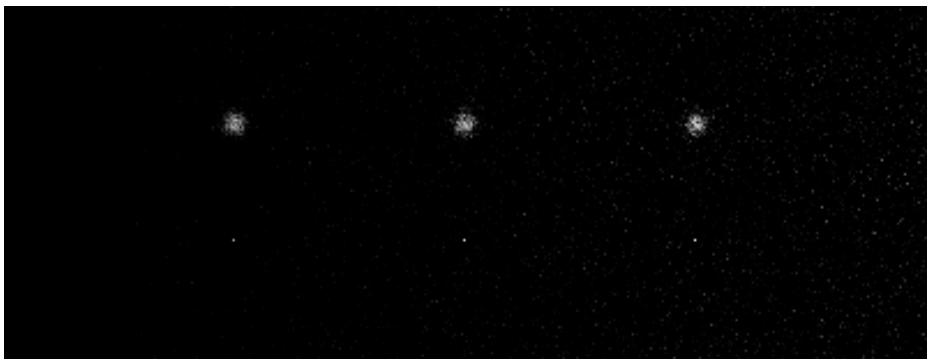

Figure 5. Adherence statistics filtered version of figure 2, with estimated background subtraction. The background noise is significantly suppressed.

Figure 4 and 5 shows the result when filtering using the adherence statistics method, without and with estimated background subtraction, respectively. As can be seen, the background noise is significantly suppressed, and hence the signal-to-noise level is greatly enhanced for both the point sources and the extended sources.

The noise suppression is indicated in figure 6, where plots of the intensities of the upper half of the images of figure 3 (left) and 5 (right) are plotted on top of each other. There is an approximate 4-10-fold background noise suppression without significant suppression of the source peak power, although the edges and weaker parts of the source signal is trimmed somewhat. For the point sources, the effect is also quite significant, as seen in figure 5. The point source intensity is virtually unchanged, and a similar 4-10-fold background noise suppression is observed.

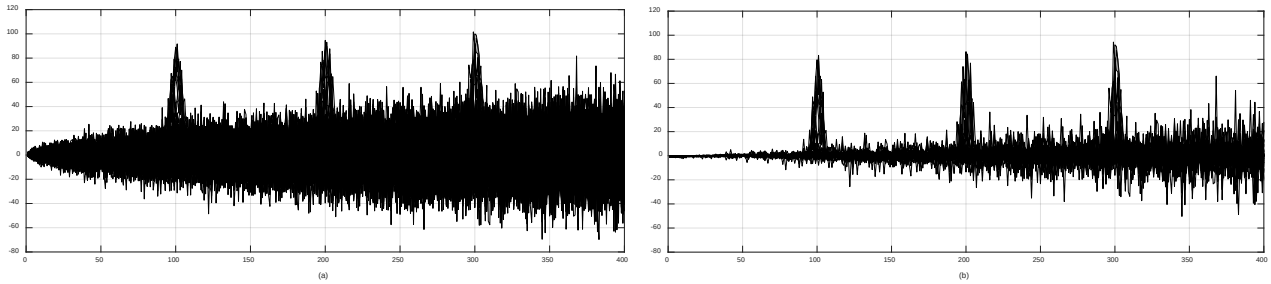

Figure 6. Background subtracted overlapping plots showing the suppression of background noise without significant suppression of the extended source signals. (a) Original source plot and (b) filtered source plot.

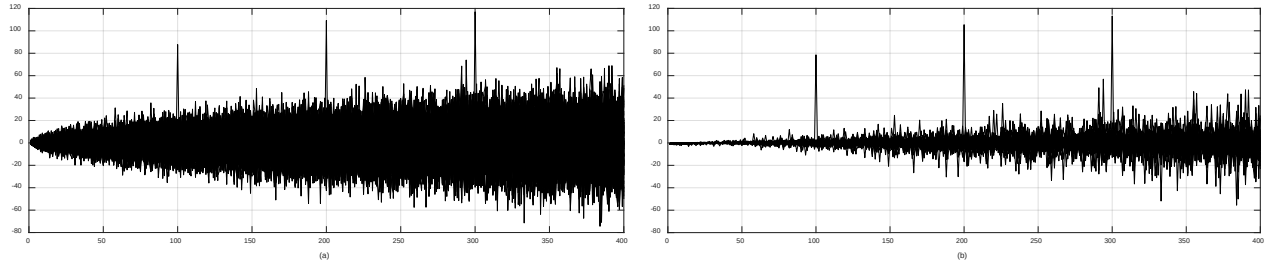

Figure 7. Background subtracted overlapping plots showing the suppression of background noise without significant suppression of the point source signals. (a) Original source plot and (b) filtered source plot.

4. DISCUSSION

The suppression of background noise here is made based on a built-in flexible adjustment of a filter that recognizes the background level and statistics. It uses the assumption that the background signal is a representation of a statistical distribution, and is hence able to replace individual samples by a more stable value representing the statistical distribution, opening up the possibility to reduce the background level below fundamental noise limits. If the signal level does not resemble the values of the statistical representation of the background, it is disconnected from this distribution, and the value neither influences nor is influenced by neighboring signals. In this way, signals above the noise level can remain relatively unchanged despite severe background suppression. Another strength of the method is its ability to adapt locally, both in terms of level as in terms of spread. Although this enables it to go below the fundamental noise level limits, it should be stressed that the method does not bring out signals buried in the background. If a signal is considered to be part of the background, it will be suppressed in the same way as the background level. Conversely, unusually strong noise spikes are not suppressed in the same way as the rest of the background noise.

In the rocket plume imaging, the signatures have for most practical purposes a positive contrast. The filtering process presented here do, however, work equally well for negative contrast signatures, as indicated in figure 1. It is even possible to suppress noise while maintaining both positive and negative contrast targets simultaneously, hence making it useful for a broad range of applications.

5. CONCLUSION

A new filtering method based on a more loosely connected statistical model using the notion of adherence has been presented and shown to enhance images with relatively low signal to noise ratios. This bears some resemblance to low pass filtering, but is still able to preserve small, even point-like objects well in the image, while smoothing out the background, thus being well adapted to image filtering of several scenes such as remote rocket plumes. Its adaptive nature and the possibility to preserve both negative and positive contrast targets with both small and large sizes simultaneously while significantly reducing background clutter or noise makes is an interesting contender for a broad range of filtering applications.

ACKNOWLEDGEMENTS

The author would like to thank Mrs. Ingebjørg Kåsen and Prof. Torbjørn Skauli for helpful discussions during the development of this work.

REFERENCES

[1] OFIL Ltd web site: Innovative UV technology for surveillance, homeland security & civil applications, 10. March 2017, http://www.sbuv.com/.

[2] Morris, P. A., Aspden, R. S., Bell, J. E. C., Boyd, R. W. and Padgett, M. J., "Imaging with a small number of photons," Nature Comm. 6, 5913 (2015).

[3] Levanon, N., [Radar Principles], John Wiley & Sons, New York, 247-267 (1988).